

## CLAIMS

What is claimed is:

1 1. A method for generating a speech recognition database comprising:  
2 generating a latent semantic analysis (LSA) space from a training corpus of  
3 documents representative of a language;  
4 receiving a new document that represents a change in the language; and  
5 adapting the LSA space to reflect the change in the language.

1 2. The method of claim 1, wherein adapting the LSA space to reflect the change in  
2 the language comprises transforming the LSA space to take into account the new  
3 document's influence on the LSA space without re-computing the LSA space.

1 3. The method of claim 1, wherein transforming the LSA space comprises:  
2 obtaining a training document vector that characterizes a semantic position of the  
3 training document within the LSA space;  
4 computing a new document vector that characterizes a semantic position of the  
5 new document within the LSA space;  
6 deriving a document vector transformation matrix; and  
7 applying the document vector transformation matrix to the training document  
8 vector and the new document vector to shift a position of each document vector in the  
9 LSA space, where the shift in the position reflects the change in the language.

1 4. The method of claim 3, further comprising:  
2       obtaining a training word vector that characterizes a semantic position of the  
3 training word within the LSA space;  
4       computing a new word vector that characterizes a semantic position of the new  
5 word within the LSA space;  
6       deriving a word vector transformation matrix; and  
7       applying the word vector transformation matrix to the training word vector and  
8 the new word vector to shift a position of each word vector in the LSA space, where the  
9 shift in the position reflects the change in the language.

1 5. The method of claim 4, wherein:  
2       the training document vector is  $VS$ , where  $VS$  is computed from a right singular  
3 matrix  $V$  and a diagonal matrix  $S$ , each of which was obtained from a previous singular  
4 value decomposition (SVD) of a training word-document matrix constructed during the  
5 generation of the LSA space, the training word-document matrix representing the extent  
6 to which each of the words appears in each of the documents of the training corpus;  
7       the new document vector  $ZS$ , where  $ZS$  is computed from the diagonal matrix  $S$   
8 and an extension matrix  $Z$ , wherein  $Z$  is an extension of the right singular matrix  
9  $V$  obtained by folding in a new word-document matrix, the new word-document matrix  
10 representing the extent to which a new word appears in the new document; and  
11       the document vector transformation matrix is  $J$ , wherein  $J$  is obtained from a  
12 Choleski decomposition of a matrix derived from an extension matrix  $Y$ , wherein  $Y$  is an

13 extension of a left singular matrix  $U$  obtained by folding in the new word-document  
14 matrix, and wherein  $U$  was obtained from the previous SVD of the training word-  
15 document matrix constructed during the generation of the LSA space.

1 6. The method of claim 5, wherein:

2 the training word vector is  $US$ , wherein  $US$  is computed from the left singular

3 matrix  $U$  and the diagonal matrix  $S$ ;

4 the new word vector is  $YS$ , wherein  $YS$  is computed from the diagonal matrix  $S$

5 and the extension matrix  $Y$ ; and

6 the word vector transformation matrix is  $K$ , wherein  $K$  is obtained from a

7 Choleski decomposition of a matrix derived from the extension matrix  $Z$ .

1 7. The method of claim 6, wherein transforming the LSA space comprises applying  
2 the document vector transformation matrix and the word vector transformation matrix  
3 simultaneously.

1 8. The method of claim 6, wherein when the new document matrix contains more

2 new documents than new words, then transforming the LSA space comprises:

3 applying the word vector transformation matrix  $K$ , first; and

4 applying the document vector transformation matrix  $J$  second, wherein the

5 extension matrix  $Y$  is not obtained by folding in the new word-document matrix, but is

6 rather derived from the extension matrix  $Z$ .

1 9. The method of claim 6, wherein when the new document matrix contains more  
2 new words than new documents, then transforming the LSA space comprises:  
3 applying the document vector transformation matrix  $J$  first; and  
4 applying the word vector transformation matrix  $K$  second, wherein the extension  
5 matrix  $Z$  is not obtained by folding in the new word-document matrix, but is rather  
6 derived from the extension matrix  $Y$ .

1 10. The method of claim 1, wherein the change in the language is a change in the  
2 language's domain.

1 11. The method of claim 1, wherein the change in the language is a change in the  
2 language's style.

1 12. A computer-readable medium having executable instructions to cause a computer  
2 to perform a method for generating a speech recognition database comprising:  
3 generating a latent semantic analysis (LSA) space from a training corpus of  
4 documents representative of a language;  
5 receiving a new document that represents a change in the language; and  
6 adapting the LSA space to reflect the change in the language.

1 13. The computer-readable medium of claim 12, wherein adapting the LSA space to  
2 reflect the change in the language further comprises transforming the LSA space to take

3 into account the new document's influence on the LSA space without re-computing the  
4 LSA space.

1 14. The computer-readable medium of claim 13, wherein transforming the LSA space  
2 further comprises:

3 obtaining a training document vector that characterizes a semantic position of the  
4 training document within the LSA space;

5 computing a new document vector that characterizes a semantic position of the  
6 new document within the LSA space;

7 deriving a document vector transformation matrix; and

8 applying the document vector transformation matrix to the training document  
9 vector and the new document vector to shift a position of each document vector in the  
10 LSA space, where the shift in the position reflects the change in the language.

1 15. The computer-readable medium of claim 14, wherein transforming the LSA space  
2 further comprises:

3 obtaining a training word vector that characterizes a semantic position of the  
4 training word within the LSA space;

5 computing a new word vector that characterizes a semantic position of the new  
6 word within the LSA space;

7 deriving a word vector transformation matrix; and

8 applying the word vector transformation matrix to the training word vector and  
9 the new word vector to shift a position of each word vector in the LSA space, where the  
10 shift in the position reflects the change in the language.

1 16. The computer-readable medium of claim 14, wherein:  
2 the training document vector is  $VS$  where  $VS$  is computed from a right singular  
3 matrix  $V$  and a diagonal matrix  $S$ , each of which was obtained from a previous singular  
4 value decomposition (SVD) of a training word-document matrix constructed during the  
5 generation of the LSA space, the training word-document matrix representing the extent  
6 to which each of the words appears in each of the documents of the training corpus;  
7 the new document vector is  $ZS$  where  $ZS$  is computed from the diagonal matrix  
8  $S$  and an extension matrix  $Z$ , wherein  $Z$  is an extension of the right singular matrix  
9  $V$  obtained by folding in a new word-document matrix, the new word-document matrix  
10 representing the extent to which a new word appears in the new document; and  
11 the document vector transformation matrix is  $J$ , wherein  $J$  is obtained from a  
12 Choleski decomposition of a matrix derived from an extension matrix  $Y$ , wherein  $Y$  is an  
13 extension of a left singular matrix  $U$  obtained by folding in the new word-document  
14 matrix, and wherein  $U$  was obtained from the previous SVD of the training word-  
15 document matrix constructed during the generation of the LSA space.

1 17. The computer-readable medium of claim 16, wherein:

2 the training word vector is  $US$ , wherein  $US$  is computed from the left singular  
3 matrix  $U$  and the diagonal matrix  $S$ ;  
4 the new word vector is  $YS$ , wherein  $YS$  is computed from the diagonal matrix  $S$   
5 and the extension matrix  $Y$ ; and  
6 the word vector transformation matrix is  $K$ , wherein  $K$  is obtained from a  
7 Choleski decomposition of a matrix derived from the extension matrix  $Z$ .

1 18. The computer-readable medium of claim 17, wherein transforming the LSA space  
2 further comprises applying the document vector transformation matrix and the word  
3 vector transformation matrix simultaneously.

1 19. The computer-readable medium of claim 17, wherein, when the new document  
2 matrix contains more new documents than new words, transforming the LSA space  
3 further comprises:

4 applying the word vector transformation matrix  $K$ , first; and  
5 applying the document vector transformation matrix is  $J$  second, wherein the  
6 extension matrix  $Y$  is not obtained by folding in the new word-document matrix, but is  
7 rather derived from the extension matrix  $Z$ .

1 20. The computer-readable medium of claim 17, wherein, when the new document  
2 matrix contains more new words than new documents, transforming the LSA space  
3 comprises:

4 applying the document vector transformation matrix  $J$  first; and  
5 applying the word vector transformation matrix  $K$  second, wherein the extension  
6 matrix  $Z$  is not obtained by folding in the new word-document matrix, but is rather  
7 derived from the extension matrix  $Y$ .

1 21. The computer-readable medium of claim 12, wherein the change in the language  
2 is a change in the language's domain.

1 22. The computer-readable medium of claim 12, wherein the change in the language  
2 is a change in the language's style.

1 23. An apparatus for generating a speech recognition database, the apparatus  
2 comprising:  
3 a latent semantic analysis (LSA) space generator to generate an LSA space from a  
4 training corpus of documents representative of a language;  
5 a document receiver to receive a new document that represents a change in the  
6 language; and  
7 an LSA space adapter to adapt the LSA space to reflect the change in the  
8 language.

1 24. The apparatus of claim 23, wherein LSA space adapter transforms the LSA space  
2 to take into account the new document's influence on the LSA space without re-  
3 computing the LSA space.

1 25. The apparatus of claim 23, wherein the LSA space adapter transforms the LSA  
2 space by:

3 obtaining a training document vector that characterizes a semantic position of the  
4 training document within the LSA space;

5 computing a new document vector that characterizes a semantic position of the  
6 new document within the LSA space;

7 deriving a document vector transformation matrix; and

8 applying the document vector transformation matrix to the training document  
9 vector and the new document vector to shift a position of each document vector in the  
10 LSA space, where the shift in the position reflects the change in the language.

1 26. The apparatus of claim 25, wherein the LSA space adapter further transforms the  
2 LSA space by:

3 obtaining a training word vector that characterizes a semantic position of the  
4 training word within the LSA space;

5 computing a new word vector that characterizes a semantic position of the new  
6 word within the LSA space;

7 deriving a word vector transformation matrix; and

8 applying the word vector transformation matrix to the training word vector and  
9 the new word vector to shift a position of each word vector in the LSA space, where the  
10 shift in the position reflects the change in the language.

1 27. The apparatus of claim 26, wherein:  
2 the training document vector is  $VS$ , where  $VS$  is computed from a right singular  
3 matrix  $V$  and a diagonal matrix  $S$ , each of which was obtained from a previous singular  
4 value decomposition (SVD) of a training word-document matrix constructed during the  
5 generation of the LSA space, the training word-document matrix representing the extent  
6 to which each of the words appears in each of the documents of the training corpus;  
7 the new document vector  $ZS$ , where  $ZS$  is computed from the diagonal matrix  $S$   
8 and an extension matrix  $Z$ , wherein  $Z$  is an extension of the right singular matrix  
9  $V$  obtained by folding in a new word-document matrix, the new word-document matrix  
10 representing the extent to which a new word appears in the new document; and  
11 the document vector transformation matrix is  $J$ , wherein  $J$  is obtained from a  
12 Choleski decomposition of a matrix derived from an extension matrix  $Y$ , wherein  $Y$  is an  
13 extension of a left singular matrix  $U$  obtained by folding in the new word-document  
14 matrix, and wherein  $U$  was obtained from the previous SVD of the training word-  
15 document matrix constructed during the generation of the LSA space.

1 28. The apparatus of claim 26, wherein:

2 the training word vector is  $US$ , where  $US$  is computed from a left singular matrix  
3  $U$  and the diagonal matrix  $S$ ;  
4 the new word vector is  $YS$ , where  $YS$  is computed from the diagonal matrix  $S$   
5 and the extension matrix  $Y$ ; and  
6 the word vector transformation matrix is  $K$ , wherein  $K$  is obtained from a  
7 Choleski decomposition of a matrix derived from the extension matrix  $Z$ .

1 29. The apparatus of claim 26, wherein the LSA space adapter transforms the LSA  
2 space by applying the document vector transformation matrix and the word vector  
3 transformation matrix simultaneously.

1 30. The apparatus of claim 26, wherein when the new document matrix contains more  
2 new documents than new words, then the LSA space adapter transforms space by:  
3 applying the word vector transformation matrix  $K$ , first; and  
4 applying the document vector transformation matrix is  $J$  second, wherein the  
5 extension matrix  $Y$  is not obtained by folding in the new word-document matrix, but is  
6 rather derived from the extension matrix  $Z$ .

1 31. The apparatus of claim 26, wherein when the new document matrix contains more  
2 new words than new documents, then the LSA space adapter transforms the LSA space  
3 by:  
4 applying the document vector transformation matrix  $J$  first; and

5 applying the word vector transformation matrix  $K$  second, wherein the extension  
6 matrix  $Z$  is not obtained by folding in the new word-document matrix, but is rather  
7 derived from the extension matrix  $Y$ .

1 32. The apparatus of claim 23, wherein the change in the language is a change in the  
2 language's domain.

1 33. The apparatus of claim 23, wherein the change in the language is a change in the  
2 language's style.

1 34. An apparatus for recognizing speech, the apparatus comprising:  
2 means for recognizing an audio input as a new document; and  
3 means for processing the new document using latent semantic adaptation; and  
4 means, coupled to the means for processing, for semantically inferring from a  
5 vector representation of the new document which of a plurality of known words and  
6 known documents correlate to the new document.

1 35. The apparatus of claim 34, wherein the means for processing the sequence of  
2 words and documents using latent semantic adaptation comprises:  
3 means for generating a latent semantic analysis (LSA) space from a training  
4 corpus of documents representative of a language;

5 means for receiving the new document that represents a change in the language;

6 and

7 means for adapting the LSA space to reflect the change in the language.

1 36. The apparatus of claim 34, wherein the means for adapting the LSA space to  
2 reflect the change in the language comprises a means for transforming the LSA space to  
3 take into account the new document's influence on the LSA space without re-computing  
4 the LSA space.

1 37. The apparatus of claim 34, wherein the means for transforming the LSA space  
2 comprises:

3 means for obtaining a training document vector that characterizes a semantic  
4 position of the training document within the LSA space;

5 means for computing a new document vector that characterizes a semantic  
6 position of the new document within the LSA space;  
7 means for deriving a document vector transformation matrix; and  
8 means for applying the document vector transformation matrix to the training  
9 document vector and the new document vector to shift a position of each document vector  
10 in the LSA space, where the shift in the position reflects the change in the language.

1 38. The apparatus of claim 37, wherein the means for transforming the LSA space  
2 further comprises:

3 means for obtaining a training word vector that characterizes a semantic position  
4 of the training word within the LSA space;  
5 means for computing a new word vector that characterizes a semantic position of  
6 the new word within the LSA space;  
7 means for deriving a word vector transformation matrix; and  
8 means for applying the word vector transformation matrix to the training word  
9 vector and the new word vector to shift a position of each word vector in the LSA space,  
10 where the shift in the position reflects the change in the language.

1 39. The apparatus of claim 38, wherein:  
2 the training document vector is  $VS$ , where  $VS$  is computed from a right singular  
3 matrix  $V$  and a diagonal matrix  $S$ , each of which was obtained from a previous singular  
4 value decomposition (SVD) of a training word-document matrix constructed during the  
5 generation of the LSA space, the training word-document matrix representing the extent  
6 to which each of the words appears in each of the documents of the training corpus;  
7 the new document vector  $ZS$ , where  $ZS$  is computed from the diagonal matrix  
8  $S$  and an extension matrix  $Z$ , wherein  $Z$  is an extension of the right singular matrix  
9  $V$  obtained by folding in a new word-document matrix, the new word-document matrix  
10 representing the extent to which a new word appears in the new document; and  
11 the document vector transformation matrix is  $J$ , wherein  $J$  is obtained from a  
12 Choleski decomposition of a matrix derived from an extension matrix  $Y$ , wherein  $Y$  is an  
13 extension of a left singular matrix  $U$  obtained by folding in the new word-document

14 matrix, and wherein  $U$  was obtained from the previous SVD of the training word-  
15 document matrix constructed during the generation of the LSA space.

1 40. The apparatus of claim 39, wherein:

2 the training word vector is  $US$ , wherein  $US$  is computed from the left singular  
3 matrix  $U$  and the diagonal matrix  $S$ ;

4 the new word vector is  $YS$ , where  $YS$  is computed from the the diagonal matrix  
5  $S$  and the extension matrix  $Y$ ; and

6 the word vector transformation matrix is  $K$ , wherein  $K$  is obtained from a  
7 Choleski decomposition of a matrix derived from the extension matrix  $Z$ .

1 41. The apparatus of claim 37, wherein the means for transforming the LSA space  
2 further comprises means for applying the document vector transformation matrix and the  
3 word vector transformation matrix simultaneously.

1 42. The apparatus of claim 37, wherein when the new document matrix contains more  
2 new documents than new words, then the means for transforming the LSA space further  
3 comprises:

4 means for applying the word vector transformation matrix  $K$ , first; and

5 means for applying the document vector transformation matrix  $J$  second, wherein  
6 the means for obtaining the extension matrix  $Y$  is not by folding in the new word-

7 document matrix, but is rather by deriving extension matrix  $Y$  from the extension matrix  
8  $Z$ .

1 43. The apparatus of claim 37, wherein when the new document matrix contains more  
2 new words than new documents, then the means for transforming the LSA space further  
3 comprises:

4 means for applying the document vector transformation matrix  $J$  first; and  
5 means for applying the word vector transformation matrix  $K$  second, wherein the  
6 means for obtaining the extension matrix  $Z$  is not by folding in the new word-document  
7 matrix, but is rather by deriving the extension matrix  $Z$  from the extension matrix  $Y$ .

1 44. The apparatus of claim 35, wherein the change in the language is a change in the  
2 language's domain.

1 45. The apparatus of claim 35, wherein the change in the language is a change in the  
2 language's style.

1 46. An system for processing speech, the system comprising:  
2 a speech recognition database comprising a latent semantic analysis (LSA) space  
3 generated from a training corpus of documents representative of a language;  
4 an input receiver to receive a new document that represents a change in the  
5 language; and

6 a processing system to adapt the LSA space to reflect the change in the language.

1 47. The system of claim 46, wherein the processing system adapts the LSA space by  
2 transforming the LSA space to take into account the new document's influence on the  
3 LSA space without re-computing the LSA space.

1 48. The system of claim 46, wherein the processing system transforms the LSA space  
2 by:

3 obtaining a training document vector that characterizes a semantic position of the  
4 training document within the LSA space;

5 computing a new document vector that characterizes a semantic position of the  
6 new document within the LSA space;

7 deriving a document vector transformation matrix; and

8 applying the document vector transformation matrix to the training document  
9 vector and the new document vector to shift a position of each document vector in the  
10 LSA space, where the shift in the position reflects the change in the language.

1 49. The system of claim 48, wherein the processing system further transforms the  
2 LSA space by:

3 obtaining a training word vector that characterizes a semantic position of the  
4 training word within the LSA space;

5 computing a new word vector that characterizes a semantic position of the new  
6 word within the LSA space;

7 deriving a word vector transformation matrix; and  
8 applying the word vector transformation matrix to the training word vector and  
9 the new word vector to shift a position of each word vector in the LSA space, where the  
10 shift in the position reflects the change in the language.

1 50. The system of claim 49, wherein:  
2 the training document vector is  $VS$ , where  $VS$  is computed from a right singular  
3 matrix  $V$  and a diagonal matrix  $S$ , each of which was obtained from a previous singular  
4 value decomposition (SVD) of a training word-document matrix constructed during the  
5 generation of the LSA space, the training word-document matrix representing the extent  
6 to which each of the words appears in each of the documents of the training corpus;  
7 the new document vector  $ZS$ , where  $ZS$  is computed from the diagonal matrix  $S$   
8 and an extension matrix  $Z$ , wherein  $Z$  is an extension of the right singular matrix  
9  $V$  obtained by folding in a new word-document matrix, the new word-document matrix  
10 representing the extent to which a new word appears in the new document; and  
11 the document vector transformation matrix is  $J$ , wherein  $J$  is obtained from a  
12 Choleski decomposition of a matrix derived from an extension matrix  $Y$ , wherein  $Y$  is an  
13 extension of a left singular matrix  $U$  obtained by folding in the new word-document  
14 matrix, and wherein  $U$  was obtained from the previous SVD of the training word-  
15 document matrix constructed during the generation of the LSA space.

1 51. The system of claim 50, wherein:

2 the training word vector is  $US$ , where  $US$  is computed from a left singular matrix  
3  $U$  and the diagonal matrix  $S$ ;  
4 the new word vector is  $YS$ , wherein  $YS$  is computed from the diagonal matrix  $S$   
5 and the extension matrix  $Y$ ; and  
6 the word vector transformation matrix is  $K$ , wherein  $K$  is obtained from a  
7 Choleski decomposition of a matrix derived from the extension matrix  $Z$ .

1 52. The system of claim 50, wherein the processing system transforms the LSA space  
2 by applying the document vector transformation matrix and the word vector  
3 transformation matrix simultaneously.

1 53. The system of claim 50, wherein when the new document matrix contains more  
2 new documents than new words, then the processing system transforms space by:  
3 applying the word vector transformation matrix  $K$ , first; and  
4 applying the document vector transformation matrix is  $J$  second, wherein the  
5 extension matrix  $Y$  is not obtained by folding in the new word-document matrix, but is  
6 rather derived from the extension matrix  $Z$ .

1 54. The system of claim 50, wherein when the new document matrix contains more  
2 new words than new documents, then the processing system transforms the LSA space  
3 by:  
4 applying the document vector transformation matrix  $J$  first; and

5 applying the word vector transformation matrix  $K$  second, wherein the extension  
6 matrix  $Z$  is not obtained by folding in the new word-document matrix, but is rather  
7 derived from the extension matrix  $Y$ .

1 55. The system of claim 46, wherein the change in the language is a change in the  
2 language's domain.

1 56. The system of claim 46, wherein the change in the language is a change in the  
2 language's style.

JUN 20 2018 2:20 PM  
U.S. PATENT AND TRADEMARK OFFICE